

The research program of the Center for Economic Studies produces a wide range of theoretical and empirical economics analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of research papers. The purpose of the Discussion Papers is to circulate intermediate and final results of this research among interested readers within and outside the Census Bureau. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Room 1587, FB 3, U.S. Bureau of the Census, Washington, DC 20233-6300, (301-763-2065).

**ANALYTIC USE OF ECONOMIC MICRODATA;  
A MODEL FOR RESEARCHER ACCESS WITH  
CONFIDENTIALITY PROTECTION**

By

Robert H. McGuckin  
Chief, Center for Economic Studies  
U.S. Bureau of the Census

CES 92-8 August 1992

### Abstract

A primary responsibility of the Center for Economic Studies (CES) of the U.S. Bureau of the Census is to facilitate researcher access to confidential economic microdata files. Benefits from this program accrue not only to policy makers--there is a growing awareness of the importance of microdata for analyzing both the descriptive and welfare implications of regulatory and environmental changes--but also and importantly to the statistical agencies themselves. In fact, there is substantial recent literature arguing for the proposition that the largest single improvement that the U.S. statistical system could make is to improve its analytic capabilities.

In this paper I briefly discuss these benefits to greater access for analytical work and ways to achieve them. Due to the nature of business data, public use databases and masking technologies are not available as vehicles for releasing useful microdata files. I conclude that a combination of outside and inside research programs, carefully coordinated and integrated is the best model for ensuring that statistical agencies reap the gains from analytic data users. For the United States, at least, this is fortuitous with respect to justifying access since any direct research with confidential data by outsiders must have a "statistical purpose".

Until the advent of CES, it was virtually impossible for researchers to work with the economic microdata collected by the various economic censuses. While the CES program is quite large, as it now stands, researchers, or their representatives, must come to the Census Bureau in Washington, D.C. to access the data. The success of the program has led to increasing demands for data access in facilities outside of the Washington, D.C. area. Two options are considered: 1) Establish Census Bureau facilities in various universities or similar nonprofit research facilities and 2) Develop CES regional operations in existing Census Bureau regional offices.

Keywords: Analytic Research, Public Use, Confidentiality, Microdata Access

\* This paper was prepared for presentation at the Eurostat sponsored International Seminar on Statistical Confidentiality, September, 1992 in Dublin, Ireland. I want to acknowledge helpful comments from James D. Adams, William Butz, Mark Doms, Jerry Gates, John Haltiwanger, Barbara J. Katz, M. Catherine Miller, Sang Nguyen, Ariel Pakes, Arnie Reznek, and Charles A. Waite and excellent secretarial assistance from Rebecca Turner. Some of the material in Section III closely parallels a panel report being prepared for the National Science Foundation. An initial version of the report was drafted by Kenneth Troske. This paper does not necessarily reflect the views of the U.S. Census Bureau.

## **I. Introduction**

One of the primary responsibilities of the Center for Economic Studies (CES) of the U.S. Census Bureau (Bureau) is to develop programs and procedures for enhancing researcher access to confidential economic microdata files. This part of its mission is a direct outgrowth of its primary task that is to bring an analytic user's perspective and insight to the economic programs area of the Bureau. In this paper I discuss the importance of microdata access to statistical agencies and researchers, describe how CES has increased researcher access, and suggest plans to expand the program to remote locations.

The paper is organized into six sections. In Section II, analytic research is described and its importance outlined. In Section III, I discuss the possibilities for analytic research with microdata, given the legal requirements surrounding data collection in the United States (U.S.). In Section IV, I discuss the history of analytic research at the Bureau, and the research associates program begun at CES in 1986. This program provides researcher access to economic microdata. It consists of a combination of outside and inside research that seeks to satisfy the needs of both researchers and statistical agencies, while ensuring confidentiality protection. In Section V, I consider the pros and cons of establishing research facilities outside the Washington, D.C. area. Two possible options for regional facilities are considered: 1) Establish Bureau facilities at universities or nonprofit research organizations. 2) Place the data in the Bureau's regional offices. Section VI concludes the paper.

## **II. The Importance of Analytic Research and Microdata**

It is useful to outline the distinction between analytic research and the kind of methodological research typically done at statistical agencies. The line between the two is not sharp, and I am trying only to provide general guidance and context for what follows. The differences are illustrated by

contrasting the type of questions that would most likely be asked of the methodologist and the analytic researcher. A methodologist might be given a budget and asked to develop a minimum variance estimate for industry output. This is a simple version of a typical survey design problem that would include, among other things, universe and sample unit definition. In contrast, the analytical user would more likely be interested in whether industry output is an appropriate measure for a particular economic or policy issue. For example, the issue might require a measure of the importance of an industry. In this case, the question is model-based, and the answer depends on the concept of "importance" -- important in what sense: job creation, size, political power, or national defense? A related question of concern to both the methodologist and the analytic researcher is whether or not measures of importance such as "industry output" perform well in particular models or policy analyses.

The analytic researcher usually approaches analysis from the perspective of an optimization model based on the objectives and constraints facing individual agents (e.g. firms, plants and individuals.) Analysis focuses on basic relationships describing the agent's behavior. Measurement errors generated when these behavioral relationships are specified at the aggregate level based on representative firm models are emphasized below.

#### **Analytic Research With Microdata**

In order to illustrate the scope and importance of analytic research, let us note that most published data reported by the U.S. statistical system are aggregations that reduce the myriad of individual detail to manageable proportions and provide confidentiality protection. Unfortunately, information is lost or distorted in this aggregation process. For some problems, this loss of detail may not matter: The phenomena under study may be sufficiently understood without reference to the underlying microdata. Without analysis of the microdata, however, it is virtually impossible to

evaluate the extent of any aggregation error. Moreover, perfectly acceptable aggregate measures at one point in time may be misleading at another point in time because the economy is constantly changing. Aggregation must therefore be approached with substantial caution and must be continually reevaluated using microdata.

It is difficult to provide a general demonstration of the importance of aggregation errors because the extent to which aggregation bias is present is model-specific. That is, the implications of the error depend on the application or use of the data. In earlier work, McGuckin (1990), I argued that the homogeneity of establishment behavior assumed in empirical studies based on aggregate data is not evident in the detailed data. Recent empirical work, Baily, Hulten, and Campbell (1991), Davis and Haltiwanger (1990), Dunne (1991) and (1992), Dunne and Roberts (1991) and (1992), Streitwieser (1991), and Troske (1991), extends this evidence by showing a striking degree of heterogeneity in the levels and movements of variables such as productivity, employment, growth, output, product structure, and investment among firms located in similar markets, industries, and cohorts.

A legitimate response to this evidence is that it is not sufficient to invalidate the use of aggregate data. Even if the behavior of the individual units to be aggregated is idiosyncratic, under certain conditions the use of aggregate variables introduces negligible bias in the estimated relationship. Unfortunately, as a long line of economic research has demonstrated, these conditions are quite restrictive. Thus, there are good reasons to believe that even if one is interested only in aggregate responses to alternative policies (such as the effect of changes in pollution regulation, defense reductions on employment in a sector, or tariff increases), aggregate industry responses will not be captured by a simple linear function of an average or representative firm because the response of individual firms to changes are very different. Industry responses are usually a weighted average of individual responses, and the weights can change over time.

In addition to variations in responses of existing firms, firm entry and exit in response to changes leads to aggregate industry responses that are not simple linear functions of the representative firm. Thus, any analysis of the aggregate effects of a policy or environmental change on a market in which there is entry and exit must incorporate not only the distribution of the response of the market incumbents, but also an analysis of the entry and exit that the policy or environmental change induces.<sup>1</sup>

The problem with exclusive use of aggregate statistics is not simply one of inferior estimates of economic relationships such as the elasticities of a production function, inventory adjustment coefficients, or wage equation parameters. With aggregate data alone, it is impossible to examine the differential effects of policies on the entities classified within the aggregate. Examining individual changes is necessary if particular components of an aggregate movement are significant. For example, regulatory changes may force firms to substitute away from labor and towards capital in production. Work by Davis and Haltiwanger (1990) at CES shows that the variation in new job creation and destruction across plants is primarily intra-industry: Almost all of the variance is within time periods, within industry variance. Thus, the effect of increased regulation on employment will depend on the detailed characteristics of an industry and cannot be captured by a representative or average industry response.

So far we have primarily considered why microdata access is important for analytic researchers and policy analysts. We now relate these to statistical agency needs. Several benefits are relatively easy to specify. First, when analysts work with the detailed microdata, that data quality can be evaluated in several new ways. Since researchers often use the data in ways that are quite different from the original data collection objective, the

---

<sup>1</sup> Failure to do so introduces selection biases to the analysis. For example, recent work at CES by Olley and Pakes (1991) demonstrates significant errors in aggregate estimates of productivity relationships in telecommunications, an industry with substantial entry and exit.

data are subjected to a variety of edits and checks that are not likely to be made in regular processing. This provides important feedback to the agency. Similarly, a substantial number of CES studies have dealt with mergers and other firm organization changes. These types of projects provide, among other things, validity checks on basic sample frame information in business registers. They also highlight the importance of various types of organization changes that the agency may or may not be monitoring, for example, leveraged buyouts.

Second, the research process often generates new products. An example, one that is currently being evaluated as part of several studies at CES, is the importance of presenting establishment-age distributions in addition to size distributions. Another, mentioned earlier, is the development of new time-series statistics on job destructions and job creations based on year-to-year observations of establishments in manufacturing.

Third, analytic work is important in evaluating whether responses to surveys accurately capture the economic concept desired. For example, capacity utilization is a particularly difficult notion to pin down. One way to examine current survey responses is in terms of an analytical model of output and investment decisions.

Fourth, analytic research projects often involve cross-survey matching of statistical units. Matching across surveys for analytic purposes often highlights coverage and sample design problems. For example, one CES project found that a substantial number of minority businesses had different industry classifications depending on the particular survey instrument.

Finally, and I believe most important, statistical agencies benefit from research projects linking survey data across time. While insights are obtained in a number of areas, much of the analytic work consists, in effect, of work on the evolution of sample frames. This is because research projects with the microdata can often be characterized as examinations of the evolution of the distribution of establishments over time. This kind of research

involves considerable effort in thinking about and characterizing sample frame changes. As such, analytic research work provides important insights on the evolution of sample frames and guards against spurious changes in aggregates associated with sample frame changes. While statistical agencies pay attention to such issues, analytic research with microdata on a longitudinal basis provides important and distinct perspectives.

### **III. Research with Microdata: The Options**

The Deputy Director of the Bureau recently described Title 13, the law governing most data activities of the Bureau, as a "zero tolerance" law, Kincannon (1991). For practical purposes, this means that the microdata collected from businesses can be used only by Bureau employees. Many countries operate under similar legal strictures and generally restrict microdata access to statistical agency employees. While concern for privacy rights is the primary reason for such policies, there is a practical reason as well: The likelihood of securing cooperation and truthful survey responses is lessened when individual respondent information is publicly released or used to identify illegal activities.<sup>2</sup> Within this legal framework, microdata access can be provided to analytic users in three possible ways -- public use data files, special tabulations performed on the microdata by Bureau employees, and direct researcher access through Special Sworn Employee (SSE) status (Title 13, Section 23C, U.S.C.). The Bureau currently uses all three of these programs to satisfy the demands for researcher access to microdata.

#### **Public Use Files**

Public use microdata files consist of data for individual reporting units, such as establishments or firms, with names and other identifiers

---

<sup>2</sup> Whether people would otherwise be less truthful and for how long the data should remain confidential are open questions.

removed to prevent disclosure of confidential information.<sup>3</sup> These files have proven to be a feasible means of providing access to significant portions of the microdata collected in demographic surveys because the data are small samples from large populations and because each individual has an equal probability of being selected. Even for demographic data, small area public use microdata files (fewer than 100,000 individuals), are not released because they are thought to offer too much risk or disclosure.

Analytic work of all kinds requires direct access to the microdata. The need, however, is particularly acute with economic data since public use databases are not feasible alternatives for releasing useful microdata files. While it is possible to produce disclosure free public use files for economic data, the types of masking technologies required make the microdata virtually unusable in economic modeling. The heart of the problem is that economic analysis requires microdata at levels of detail that are characterized by cells that include very few observations and come from highly skewed populations. See Fuller (1991), McGuckin and Nguyen (1990).

### **Special Tabulations**

While valuable, special tabulations are not a complete answer to the access problem. For one thing, analytic users often question the data editing and imputation strategies used by statistical agencies in producing published statistics. It is difficult for the users to examine the validity of these techniques or adjust the data for microeconomic models without at least some "hands on" access. Moreover, even if the data are clean, specifying the models and estimation procedures that regular employees of the Bureau will execute is not as efficient for analytic researchers as hands-on access. The large number of conditional decisions that analytic researchers must make in

---

<sup>3</sup> Creation of public use files usually requires additional adjustments such as top coding or rounding.

the course of their research, and their need to make "judgment calls" makes hands-on access almost a necessity.

On the other hand, many requests for economic and statistical estimates do not involve difficult judgment calls and can be handled as special tabulations. For example, in some cases, it is possible to satisfy particular users with some form of a variance-covariance matrix. General use of this technique to create public use databases is limited, however, because of complementary disclosure requirements. In addition, the impossibility of making nonrandom extensions of the variance-covariance matrix severely limits this method of serving analytic users. See McGuckin and Nguyen (1990).

### **Special Sworn Employee (SSE) Status**

The CES program relies heavily on the extension of SSE status to analytic researchers working on-site at the Bureau. The CES has focused on analytic users because of its basic mission of research but is now beginning to expand special tabulations services. SSE researchers at CES are subject to the same legal proscriptions (possible fines of \$5000 and imprisonment for 5 years) as regular employees of the Bureau. Thus, both types of employees have similar incentives to preserve the confidentiality of data. Regular employees may more readily lose their jobs if they reveal confidential data. However, even if they do not directly lose their jobs, the loss of reputation and the denial of future access to microdata are significant deterrents to a non-staff researcher's violating confidentiality.

SSEs working at CES are immersed in the culture of confidentiality that is not only a part of the culture of the Bureau, but also is a regular part of CES operations. All researchers at CES are given periodic reminders of their responsibility to maintain the confidentiality of microdata, and all SSEs are required to sign special forms outlining CES security procedures and regulations. In addition, CES takes an active role in disclosure analysis, counseling SSEs from the beginning of their projects on the types of

tabulations that may be released and checking all research outputs carefully to ensure confidentiality protection.

The key requirement for researcher access to the microdata via SSE status is that the statistical agency must benefit from the research. Thus, it is not enough to justify access by showing an important need by analytical researchers: We also need to establish that the interests of researchers and their sponsors coincide with the research needs of the statistical agency. Frequently, in fact, the interests of analytic researcher's projects and statistical needs do coincide. This possibility has been recognized in Bureau fellowship programs, and, by and large, fellowship projects have been very useful to the Bureau. The CES program exploits the convergence of interests recognized in the fellowship program in order to provide analytic research capabilities to the economic programs area of the Bureau.

#### **IV. The CES Program**

##### **Background**

While the situation is far better than it was at the time Professor Scherer made his statement (1980), his lament retains some validity:

"The data, collected at an expense of tens of millions of dollars, lie unanalyzed in Bureau files. Though less apt to draw headlines than Congressional junkets and the overpayment of welfare recipients, this state of affairs is equally wasteful."

For many years, it was virtually impossible for researchers to gain access to the economic microdata collected in the various economic censuses and surveys. Resource limitations -- of space, fellowship money, research funds, and computer technology -- have always constrained analytic research. However, resource limitations were never the only, perhaps not even the primary, constraint on analytic research with economic microdata. A culture

that did not see analytic research as an appropriate or necessary endeavor has been pervasive in the economic program areas of American statistical agencies.<sup>4</sup>

Triplett (1991) argues that statistical managers at U.S. statistical agencies lack "knowledge of economic research and policy analysis, and the ability to communicate with analytic users of data." He argues -- correctly in my view -- that research needs for economic data have seldom been factored into the statistical planning process. Triplett also argues that the failure of statistical agencies to take sufficient account of analytic research can be traced to the strong macroeconomic bias in the Washington, D.C., policymaking environment, reinforced by a decentralized statistical system that makes it difficult for analytic users to communicate with statistical managers.<sup>5</sup> In this environment, analysis relies on economic statistical data products that consist of cross-sections of industry data or linked time-series of cross-section aggregates. Thus, analytic projects tend to pay little attention to the microdata.

### Precursor Programs

Consistent with a culture that de-emphasizes analytic research, access to the economic microdata has traditionally been limited to researchers hired by the Bureau for specific projects or those working on Joint Statistical

---

<sup>4</sup> Historically, programs of analytic subject matter research, both inside and outside the Bureau, have been stronger in the demographic programs area than in the economic programs area. This can generally be traced to two factors. First, there has been widespread availability of microdata records to demographic analytic researchers through public use files. Second, most of the analytic research resources of the economic program area were shifted to the Bureau of Economic Analysis (BEA) in the early 1970s. This political decision effectively eliminated analytic research in the economic program area until recently.

<sup>5</sup> The shift of analytic research functions away from the Bureau, coupled with steady resource cutbacks in the statistical agencies over the last 10-15 years, have reinforced the bias by increasing the costs of microdata research. In addition, the support for microdata research was much less prevalent among academic and policy economists in the 1970s and 1980s than it is today.

Agreements (JSAs). (For the most part, JSA projects have focused on statistical methodology.) There also have been special fellowship programs -- the American Statistical Association/National Science Foundation/U.S. Census Bureau (ASA/NSF/Census) Fellowship Program, and a largely unused (for lack of funding) research fellowship program designed to bring researchers from other Federal agencies to the Bureau.

The ASA/NSF/Census program began in 1978. In its early years, the program focused mainly on statistical methodology, particularly on time-series research. Later, the program undertook substantial numbers of projects involving demographic data. But it was not until 1983, one year after CES was formed, that economic researchers joined the program. Of the first 32 ASA/NSF/Census fellows from 1978 through 1986, only four used the Bureau's economic data; only two of these were analytic researchers using economic microdata. The other two projects involved time-series analysis. Since 1987, the period associated with the extended CES research associates program, there has been a dramatic increase in the number of analytic researchers working with economic microdata. Of the 26 Fellows in the program from then through 1992, including those starting in the fall of 1992, ten will have worked with economic microdata.

#### **A Note on Funding**

Aside from the issue of culture, there is an additional constraint on the expansion of analytic research: The fellowship programs can support only a limited number of researchers because project costs have been shared with outside agencies, but the researchers still require significant Bureau resources. This approach limits analytic projects to those to which the Bureau is able to allocate its limited funds. In fact, James Smith (1991) recently argued that resource constraints with the fellowship program made "it simply not feasible in the foreseeable future that the SSE model could be

applied [for researcher access] on any reasonable scale." A similar concern is expressed by Gates (1988).

Because of the scarcity of Bureau resources, if Bureau funds are the major source of research funding, then many valuable projects will not be undertaken. The CES program, while not eliminating the need for Bureau funds, provides a method for financing expanded analytic research and a mechanism for integrating this research into the Bureau's planning and development process.

### **CES Organizational Structure**

The CES research program is currently organized into six areas of research and analysis: Energy and Environment, Organization and Behavior of Firms and Markets, Labor Market Issues, Productivity and Technical Change, Entrepreneurship, and International Issues.<sup>6</sup> At least one Ph.D. economist has responsibility for developing research projects of academic quality for each of these areas, and publication in the field is a prerequisite for the economist's job advancement. Each economist oversees a micro database that is also the basis for her research. In this way the economist becomes expert in a research area and knowledgeable about the database required. The economist is then responsible for providing substantive guidance to visiting research associates and for communicating research findings to the Bureau's data collection divisions.

The research staff is supported by a computer group that procures, installs, and administers computer equipment and data. The computing group uses a distributed data processing structure with workstations and personal computers. Researchers rely on commercially available software packages that they use to manipulate the data directly. The computer group provides in-house software to retrieve data from the rest of the Bureau and develops data

---

<sup>6</sup> CES also has a program area dealing with statistical and database issues.

extraction software that are used to supply researchers with data in forms compatible with the researcher's software.

This distributed processing environment offers several advantages over the more centralized processing environments typically found with mainframes and minicomputers. Each user has enough computing power to complete her individual projects, and the level of computing power is matched to the individual's requirements. Each user is the master of her computing environment, and there is little reliance on system administrators to supply computing support except for data extraction and related software. Individual users do not adversely affect each other in a distributed environment as often as in a more centralized processing environment. Computer "down time" is limited to one or a few users when a machine fails, whereas a hardware failure with a centralized machine halts all processing for all users. Moreover, by isolating individual users and limiting each researcher's access to only the data required for her project, the computing environment improves security.

The CES research is also supported by a newly formed group with responsibilities for product development. This group includes data specialists, writers, statisticians, and analysts. These staff members are the primary supporters of CES's tabulation program; they also engage in data development research projects. They develop reports, statistical briefs, and papers designed to communicate the results of the research program to policymakers and others interested in non-technical expositions. In addition, they often provide consulting support on commercial econometric packages used by the researchers.

#### **Microdata at CES**

The data to support CES research come from the Bureau's existing surveys and censuses that underlie the aggregate published information, as well as from outside sources. The CES engages in no data collection activities itself. Databases are developed at CES by linking microdata from disparate

surveys and creating longitudinal versions of cross-section data. For example, the Longitudinal Research Database (LRD) is an unbalanced panel of establishments including all censuses of manufacturers since 1963 and all annual surveys of manufactures since 1972. The LRD contains basic information on the inputs and outputs of individual establishments. Information from numerous surveys is linked to the LRD for particular analyses. Examples include data on firm-level R&D expenditures, plant-level environmental emissions, and energy consumption. The CES is currently attempting to link demographic information on individuals in the workforce to LRD establishments.

These new and unique databases increase the value of existing data at very little cost: In effect, they represent recycling efforts. The databases support analytic research conducted at CES and also generate new data products. For example, CES now publishes a longitudinally based series of job creations and destructions. By using the new databases in research efforts, the analytical research program provides important information on the quality of the data. In turn, this leads to new data and imputation strategies, as well as improvements in survey design and in measurement of economic concepts.

#### **Development of the CES Program**

The success of the CES program is based on the willingness of economic programs management at the Bureau to (1) fund an internal program of analytical research, and (2) encourage a complementary set of outside research projects that are self-funding. While the Bureau has always conducted some internal research, most of its economic area research efforts have focused on statistical methodology. The CES program, as well as the culture change signalled by the Bureau's support of CES's analytically based program, have made it possible for CES to attract outstanding complementary projects by outside researchers.<sup>7</sup> Moreover, CES research associates bring with them non-

---

<sup>7</sup> The culture changeover is not complete, however. While analytical research has been elevated to the level of a strategic plan initiative at the Bureau, there is still resistance to the CES type of program.

Bureau funds to cover the costs of their projects. This has made it possible to overcome many of the resource constraints associated with the fellowship programs, thereby greatly expanding research access and improving economic data products.<sup>8</sup>

An internally funded program of research is extremely important to successful research access by SSEs. Without it, neither the agency nor the visiting researchers are well served. An active internal research effort ensures that the agency has the expertise to evaluate, assimilate, and communicate to the Bureau's production divisions the results of research by visiting research associates.<sup>9</sup> It also ensures that agency personnel who have enough research experience with the data can provide the advice, consultation, and related support required by the research associates.<sup>10</sup>

While not all research proposed for SSE access provides significant benefits to the Bureau, most well-conceived empirical research projects with microdata generate a bundled product -- analytic paper(s), policy guidance, and information that can be used to improve the Bureau's statistical program. Moreover, even when proposed research projects do not focus on a problem of current concern, or when the benefits to the Bureau are not specified, we have found that well-conceived empirical projects almost always have important agency benefits. In particular, analytic research is often conducted in areas of policy interest, and today's research generates demands for new statistical products in the future.

---

<sup>8</sup> I am not suggesting that a statistical agency's entire research program can be handled through the use of the SSE research associate model. While privately financed projects are generally beneficial to the statistical agency, they cannot be relied upon for all of an agency's research needs.

<sup>9</sup> Although not of direct interest here, internal expertise also enables the statistical agency to efficiently contract and monitor research done outside the regular research program and to engage in joint statistical projects in which the costs of the research are shared.

<sup>10</sup> This point is discussed further in the next section, in which I discuss the costs and benefits of regional centers.

## V. **Establishing Research Centers Away From Central Headquarters**

The increasing importance given to the microfoundations of economic models by macroeconomists, coupled with the longstanding unsatisfied needs of traditional microdata users, has led to substantial demand for CES services. The success of current CES research programs will, in all likelihood, continue to fuel demands for data access and a large number of measurement projects with significant benefits to the Bureau. Recently, however, shortage of space at central headquarters has forced CES to turn away some projects for which funds were available.<sup>11</sup> Both factors suggest that expanded CES facilities be developed.

Even though some CES expansion at headquarters is planned, it is unlikely that there will be sufficient space at headquarters to satisfy researcher demands for access or to exhaust Bureau benefits from analytic projects.<sup>12</sup> Moreover, the costs (travel, family disruption, etc.) that research associates must bear if they work at central headquarters are not trivial. Requirements to work on-site, therefore, pose serious constraints to the quality and quantity of research produced. Regional facilities provide solutions to both problems.

Current Bureau policy requires all employees, including SSEs, to use confidential data at a secure Bureau site. While special employee regulations enable CES to provide outside researchers with access to microdata, current Bureau policy requires analytic users to work only at the central headquarters offices in Washington, D.C..<sup>13</sup> Two principal issues must be addressed before

---

<sup>11</sup> The difficulties in securing sufficient space at central headquarters are a major component of the inadequate resources referred to earlier.

<sup>12</sup> As a partial indicator of the excess demand, I note that microdata research based on public use demographic data files and in-house demographic programs research exceeds by far current microdata research with economic data.

<sup>13</sup> On rare occasions, SSEs have been permitted to work away from the Bureau headquarters. None of these cases involved CES research.

a decision to expand to sites away from headquarters can be considered: security and funding. The issues of project choice and headquarters involvement in the regional centers, are closely related to funding because they involve the role of non-Bureau funding sources in the operations of the regional centers. We take up each of these issues in this section.

### **Security**

There is one major potential pitfall in expansion of research to regional sites. Public perception of lax security control at statistical agencies could increase, and this could reduce survey response rates. Unlike the situation at a national security defense facility, where products are developed directly on-site, the cooperation of the outside public is essential in statistics production and, therefore, perceptions are important. Confidential data are, however, already handled in regional offices, where there also exists a culture of security. Moreover, with CES personnel on-site, agency role models would be directly available to demonstrate and enforce proper data handling procedures. Thus, security should be relatively easy to arrange at regional statistical offices and this should alleviate the concerns about public perceptions of lax security control. In short, maintenance of security should be relatively straightforward at regional statistical offices, assuming current CES procedures are followed and the offices are staffed with CES personnel.<sup>14</sup>

I see no reason to believe similar security arrangements could not be developed at, for example, university centers. While the physical arrangements that are created are likely to be more complicated than at non-Bureau facilities, they are readily solvable. Prime issues that must be considered are the nature of the space arrangements that would be needed and how much personnel would be required -- guards, census employees, etc. -- to satisfy

---

<sup>14</sup> It could be argued that as the number of research sites expands, the possibility of leaks also increases. This same argument could be applied to any increase in the number of Bureau employees.

confidentiality standards. In principle, there is no reason that a university (or other research site) could not be designated a Bureau site. The key question for the Bureau is whether or not such site agreements would affect public perception and response rates adversely.

### **Funding**

The now discontinued ASA/NSF/Census Fellowship Program offers a model for the funding and guidance of CES regional centers. In the fellowship program model, representatives from the NSF, ASA, and the Bureau reviewed academic research proposals for quality and applicability. Funding for the fellows and ASA administrative services was provided by NSF, with research assistants and space purchased by the Bureau, and equipment funding shared by NSF and the Bureau. The Bureau was only one of a number of agencies for which NSF provided start-up funding. After NSF funding was discontinued, the Bureau instituted a smaller fellowship program that is administered and completely funded by the Bureau.

Unlike the fellowship program, outside sources would be needed for both initial and continuing funding of regional centers. A consortium of research organizations such as NSF, universities, and foundations could be formed to provide the CES centers with start-up costs financing. Such a consortium could also provide continuing funding through laboratory fees added to researchers grants, much as the NSF and foundations do now for CES research associates' projects. With the help of professional societies such as the American Economic Association (AEA) and the ASA, the consortium could also help guide the operation and projects undertaken by regional centers. This guidance would be crucial if more projects are proposed than available research capacity can handle.<sup>15</sup>

---

<sup>15</sup> Note that research capacity means more than computers and space: It also means in-house research expertise to support the associate's data needs.

One way to allocate scarce research capacity is on the basis of price. This mechanism is however, not available in a government agency. A more realistic alternative is to balance the value of the project to the Bureau against the value of the project to the funding agencies as has been done with the fellowship program. Professional peer review services have proven invaluable to the Bureau in the ASA/NSF/Census Fellowship Program, and an AEA Advisory Committee to the Bureau has provided extensive guidance to the CES program.

One might take the view that the magnitude of currently foreseeable benefits to the Bureau should be the sole determining factor. I suspect that this is not a good approach because it would leave by the wayside very important social research that might generate large benefits later, but that has little direct effect on current programs. The arguments for support of high value research, even with relatively minor current statistical agency benefits, are actually akin to those used to support basic research, which has very high, but uncertain payoffs compared to the lower, but more certain benefits of applied research.

Research into policy issues of large social concern highlights new needs for economic data. For example, examination of a model to predict inflation may suggest that new measures of capacity are necessary for accurate prediction. This kind of research into concepts and methods is the everyday focus of economic research, and such research needs to be incorporated into the statistical agency's design and development functions Triplett (1991). This is the most important contribution of analytic research because it keeps the statistical agency on the frontier of emerging data requirements. It is with this in mind that the priorities of the research community need to be incorporated into the Bureau's project choices.

#### **Headquarters Involvement in Regional Centers**

Most projects undertaken at regional offices should be chosen under procedures similar to those currently undertaken at CES and would involve CES evaluation. Disclosure analysis and data extracts would be performed best centrally. It also seems sensible to require that researchers consult with the in-house CES staff, particularly at the beginning of the project, but periodically thereafter as well. At the end of the project, a visit to headquarters is essential to ensure that documentation of the project is complete and for the researcher's consultation with CES staff and Bureau officials on the results of the project.

Thus, even with regional centers, researchers would need to spend some time at the central office. While some CES staff members would be required at the regional site for security, clerical, and computer assistance, the level of support necessary to reproduce CES services in their entirety is likely to be too expensive for a regional center to support. Moreover, the researcher needs to visit the central headquarters in order to interact with the production staff. Thus, typically the researcher initially would come to Washington, D.C., to discuss a project, and perhaps to clean data. The analysis of the data could be done nearer the researcher's home and a return trip(s) for a presentation of findings could be made later. This approach would make a researcher's dislocation costs at least manageable.

Insisting that researchers visit central headquarters ensures that the researcher works closely with agency personnel and has contact with other visiting scholars experienced with the data. Research with microdata is unlike research with published data which are often taken as given. Specifics of the sampling procedures, handling of outliers, and imputation strategies become extremely important issues in microdata research. Such factors were of primary concern to the AEA which, when a CES regional research program was first proposed, recommended that the CES focus its initial efforts on building

its research program at central headquarters.<sup>16</sup> A related important point is the importance of headquarter's involvement in making the fruits of outside research available to the Bureau.

## **VI. Concluding Comments**

In the past, there have been two problems with using SSE status as a means of providing access to analytic users. One problem, discussed in Section VI, is that even approved researchers must come to the central Washington office of the Bureau to conduct their research. The development of CES regional facilities could help alleviate this problem. In most cases, efficient operation would require a short stay at the central facility at the beginning of the project and later visits for seminars, transmission of research results, and consultations. The regional centers would be designated statistical offices staffed with at least two or three CES trained staffers.

The other problem with using SSEs as the basis for researcher access to confidential microdata is that SSE status is available only to researchers whose research benefits the Bureau -- that is, where the research satisfies a Bureau program need. Until recently, the importance of analytical research to the Bureau's economic programs was not widely recognized. The CES program is based on a growing recognition of the importance of analytical research with economic microdata in the development of quality statistical products. This recognition provides the basis for interpreting the Bureau purpose requirement to encompass research beyond narrowly focused methodology projects. The CES experience suggests that not only is a broad range of analytical research relevant, but the funds for conducting this research can be obtained from laboratory usage fees.

Finally, the evidence from CES' experience suggests that a statistical agency's analytic program of research should focus on important policy issues

---

<sup>16</sup> Over time the need for headquarters consultations on data issues may decline as documentation is developed and particular databases experience multiple users and, concomitantly, improved quality.

and emerging data needs, as well as serve as a catalyst for bringing the best researchers to the statistical agency. The expertise of the in-house staff fosters communication with leading academic researchers, and reduces the costs of research associates by helping to define necessary and feasible projects, and provides an environment that allows visiting research associates to operate efficiently. The in-house staff further ensures that the statistical agency is able to assimilate the analytic research and stay at the forefront of data needs.

## References

- Baily, Martin N., Charles Hulten and David C. Campbell "Productivity Dynamics in Manufacturing Plants," Brookings Papers on Economic Activity, Washington, D.C., 1992.
- Davis, Steven J. and John Haltiwanger "Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications," National Bureau of Economic Research Macroeconomics Annual, MIT Press, 1990.
- Dunne, Timothy "Technology Usage in U.S. Manufacturing Industries: New Evidence from the Survey of Manufacturing Technology," CES Discussion Paper Series #91-7, 1991.
- \_\_\_\_\_ "Discussion of Secondary Data Analysis," Proceedings, Census Bureau Annual Research Conference, 1991.
- Dunne, Timothy and Mark J. Roberts "The Duration of Manufacturing Employment," forthcoming, Review of Economics and Statistics, 1992.
- \_\_\_\_\_ "Costs, Demand, and Imperfect Competition as Determinants of Plant-Level Output Prices," CES Discussion Paper Series #92-5, 1992.
- Fuller, Wayne "Use of Masking Procedures for Disclosure Limitation," prepared for the Conference on Data Access Through Disclosure Limitation, Social Science Research Council, 1991.
- Gates, Gerald "Census Bureau Microdata: Providing Useful Research Data While Protecting the Anonymity of Respondents," 1988 Proceedings of the Social Statistics Section of the American Statistical Association.
- Juster, F. Thomas, Comments on "The Federal Statistical System's Response to Emerging Data Needs" by Jack E. Triplett, Journal of Economic and Social Measurement, Volume 17, 1992.
- Kincannon, C. Louis "Discussion" at American Statistical Association session "Whom Can You Trust: A Panel Discussion of Tension", 1991 Proceedings of the Social Statistics Section of the American Statistical Association, forthcoming.
- McGuckin, Robert H. "Longitudinal Economic Data at the Census Bureau: A New Data Base Yields Fresh Insights on Some Old Issues," Proceedings of the Statistics Canada Symposium on Analysis of Data in Time, 1990, pp. 243-249.
- McGuckin, Robert H. and Sang V. Nguyen "Public Use Microdata: Disclosure and Usefulness," Journal of Economic Measurement, Volume 16, Number 1, 1990, pp. 19-40.
- Olley, G. Steven and Ariel Pakes "Dynamic Behavioral Responses in Longitudinal Data Sets: Productivity in the Telecommunications Industry", Center for Economic Studies Discussion Paper 92-2, 1992.
- Scherer, F.M. Industrial Market Structure and Economic Performance, Chicago, Illinois, 1980.
- Scheuren, Fritz, Comments on "The Federal Statistical System's Response to Emerging Data Needs", Journal of Economic and Social Measurement, Volume 17, 1992.

- Singer, Burton, Comments on "The Federal Statistical System's Response to Emerging Data Needs" by Jack E. Triplett, Journal of Economic and Social Measurement, Volume 17, 1992.
- Smith, James "Data Confidentiality: A Researcher's Perspective," 1991 Proceedings of the Social Statistics Section of the American Statistical Association, forthcoming.
- Streitwieser, Mary L. "The Extent and Nature of Establishment Level Diversification in Sixteen U.S. Manufacturing Industries," Journal of Law and Economics, Part 2, November, 1991, pp. 496-551.
- Triplett, Jack "The Federal Statistical System's Response to Emerging Data Needs," Journal of Economic and Social Measurement, Volume 17, 1992.
- Troske, Kenneth "The Growth and Survival of Firms: Changes in the Firm Size Distribution in Industries," Ph.D. Dissertation, University of Chicago, Chicago, Illinois, 1992.